CMS Physics Analysis Summary

Contact: cms-pag-conveners-ewk@cern.ch 2009/08/27

Study of the ratio of W + jets to Z + jets in proton–proton collisions at $\sqrt{s} = 10$ TeV with the CMS detector at the CERN LHC

The CMS Collaboration

Abstract

We present a simulation study of the ratio of W + jets to Z + jets in pp collisions at $\sqrt{s} = 10$ TeV and demonstrate the feasibility of measurements of the ratio up to four inclusive jets with $\mathcal{O}(100)$ pb^{-1} of early CMS data at the LHC. These measurements allow early understanding of W + jets and Z + jets production at the LHC, since several detector effects and theoretical uncertainties cancel out in the ratio. They also provide a means of implementing a data-driven normalization of the W + jets and Z + jets backgrounds to important Standard Model processes (e.g. top quark) and new physics searches.

1 Introduction

Important Standard Model (SM) and new physics (NP) processes at the LHC are expected to produce final states with a vector boson (VB = W, Z) and multiple jets. The VB + jets associated production has been used at the Tevatron both as a stringent test of perturbative QCD predictions and as a means to an accurate description of backgrounds to NP [1–3]. VB + jets production at the Tevatron is an important background to top quark production; at the LHC ($\sqrt{s} = 10$ TeV) this connection is important in both directions, since W + jets production is already significantly contaminated by top quark processes for ≥ 2 jets.

Within the SM the VB $+ n$ jets cross section is $\mathcal{O}(\alpha_s^n)$. The VB $+ n$ jets over VB $+ (n + 1)$ jets yield ratio (C_{VB}) is then nearly constant as a function of n for $\mathrm{p\bar{p}}$ $\sqrt{s} = 630$ GeV and $\mathrm{p\bar{p}}$ $\sqrt{s} = 1.8$ TeV both at the parton level and in data [4–7]. QCD[1] predicts for the W+jets and Z+jets very similar accompanying final states as one would naively expect, resulting in a double ratio $C_{\mathrm{W}}/C_{\mathrm{Z}} \equiv \frac{\mathrm{W} + n \text{ jets}/\mathrm{W} + (\mathrm{n} + 1) \text{ jets}}{\mathrm{Z} + n \text{ jets}/\mathrm{Z} + (\mathrm{n} + 1) \text{ jets}}$ consistent with 1, independent of jet multiplicity.

The purpose of this analysis is: i) to measure the double ratio at different jet multiplicities at pp $\sqrt{s} = 10$ TeV and investigate to what extent the double ratio is in fact independent of jet multiplicity and ii) to provide a method of determining the W + jets absolute rate normalization in the higher jet multiplicities, given the absolute rate of Z + jets events [9]. A strong dependence of the double ratio on the number of jets could be an indication of NP [10].

In the double ratio, uncertainties that grow rapidly with n cancel [11], allowing us to contemplate performing this study at the startup of the LHC. The cancellation is predominantly due to the correlation in the jet counting uncertainties in the numerator and denominator, independent of the jet definition. Other systematics associated with the luminosity, parton distribution functions, detector acceptance and efficiencies are also expected to substantially cancel.

We focus on the LHC startup and assume $\mathcal{O}(100)$ pb^{-1} of data collected with the CMS detector [12] at a centre-of-mass energy $\sqrt{s} = 10$ TeV. We use two independent jet definitions: one based on calorimetry deposits (calo-jets) and one based on tracks (track-jets) to test the jet counting with different detector effects and to allow sampling of different parts of the phase space. We further validate the results with corrected calo-jets and particle flow jets (PF-jets) [13]. We present the simultaneous selection of $\mathrm{W}(\rightarrow \ell\nu)$ + jets and $\mathrm{Z}(\rightarrow \ell\ell)$ + jets ($\ell = \mathrm{e}, \mu$).

2 The CMS detector

A detailed description of the Compact Muon Solenoid (CMS) experiment can be found elsewhere [12]. The central feature of the CMS apparatus is a superconducting solenoid, of 6 m internal diameter. Within the field volume are the silicon pixel and strip tracker, the crystal electromagnetic calorimeter (ECAL) and the brass-scintillator hadronic calorimeter (HCAL). Muons are measured in gas chambers embedded in the iron return yoke. In addition to the barrel and endcap detectors, CMS has extensive forward calorimetry.

CMS has a two-level trigger system. The Level-1 trigger, based on custom hardware, is designed to reduce the collision rate of 40 MHz to approximately 100 kHz. The High Level Trigger (HLT) employs a set of sophisticated software algorithms that analyze the complete event information and further reduce the accepted event rate for permanent storage and analysis.

[1] Next-to-Leading Order QCD calculations are available for up to three (two) jets for W [4, 5, 7, 8] (Z [4, 5]).

3 Signal and background samples

The $W(\rightarrow \ell\nu) + n$-jets events ($\ell = e, \mu$) and $Z(\rightarrow \ell\ell) + n$-jets events ($\ell = e, \mu$) are studied with Monte Carlo simulation, using the MadGraph [14] event generator, based on a leading-order calculation of the matrix element (ME). ME calculation is performed for final states with at most four primary partons, requiring that the parton p_T exceeds 10 GeV/c. PYTHIA [15] is used for the parton shower, hadronization and the underlying event description. Parton shower matching is applied to avoid double counting of emissions in overlapping phase space regions. The MLM [16] matching algorithm with k_T clustering is used with matching threshold 15 GeV/c. The lepton pair invariant mass is required to be $m_{\ell\ell} > 50$ GeV/c^2 at the generator level. The CTEQ6L1 [17] parton distribution functions are used.

A large background component for this analysis comes from QCD multijet production. This is studied using a sample of Monte Carlo events generated with PYTHIA. Using a filter that selects electron and muon enriched multijet samples, the generation includes $b\bar{b}$, $c\bar{c}$ and decays of long-lived light mesons as sources of muons and loosely isolated hadrons or jets with an increased electromagnetic fraction as a source of electrons. The filter also requires an outgoing parton with $p_T > 20$ GeV/c. The $Z(\rightarrow \tau^+\tau^-) + n$-jets and $W(\rightarrow \tau\nu) + n$-jets events contribute to the background and are generated as part of the full $Z(\rightarrow \ell\ell) + n$-jets and $W(\rightarrow \ell\nu) + n$-jets (with $\ell = e, \mu, \tau$) samples with MadGraph and PYTHIA, using the same phase space requirements and parton shower matching settings as the signal. The $t\bar{t}$ + jets and single top backgrounds are generated with MadGraph interfaced with PYTHIA with the associated parton $p_T > 20$ GeV/c and matching threshold 30 GeV/c. While for the electron channel the γ + jets background is not considered it can be subtracted in an unbiased way as other studies also indicate. Other potential backgrounds such as diboson production are not considered since they are found to be negligible (*cf* also [18]).

4 Event reconstruction and selection

Throughout the analysis all the requirements are optimized on Monte Carlo simulation events, taking into account the correlation among the variables. The analysis path is as follows: events are filtered using an emulation of the CMS Level-1 trigger system (L1) and the offline high level trigger (HLT). Event are then reconstructed and tracks, electrons, muons, and significant energy deposits in the calorimeters are identified. The VB candidates are formed from lepton candidates in the event. Jets are formed and the events are streamed in datasets according to the n inclusive jet multiplicity ($\geq n$). An extended and unbinned maximum likelihood (ML) fit is used to determine the signal yield for each jet multiplicity and the fit results are used to measure the double ratio.

4.1 Trigger selection

The events are selected by the CMS Level-1 (L1, as emulated in the simulation) and High-Level (HLT) single electron and muon triggers with no requirement on the lepton isolation. The trigger p_T thresholds are those determined in CMS for low luminosity running ($L = 10^{32}$ cm^{-2} s^{-1}). The trigger paths used for both the W and Z selection are the HLT single 'non-isolated' muon and electron with thresholds 15 GeV/c and L1 thresholds 12 GeV/c and 10 GeV/c for electrons and muons respectively.

4.2 Lepton reconstruction and VB selection

Muons are reconstructed using an algorithm that combines the information from the muon chambers and the silicon tracker [19], and muon isolation is imposed by considering a cone around the muon defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \leq R_{\text{cone}} = 0.35$ and requiring that the sum of the p_{T} of the tracks in the cone, excluding the muon track, be less than 13% of the muon transverse momentum. Additional calorimetric isolation criteria (both in the electromagnetic calorimeter and hadronic calorimeter) are used to suppress the multijet background contribution. Electron identification is based on a standard set of criteria including various track-matching and shower shape variables in the electromagnetic calorimeter barrel and endcap regions. In addition, a tracking electron isolation criterion is applied by requiring the sum of the p_{T} of the tracks compatible with the electron vertex and $p_{\text{T}} > 1.5\,\text{GeV}/c$ around the electron candidate track in a cone of size $R_{\text{cone}} = 0.4$ to be less than 3% of the electron candidate momentum. Tight calorimetric isolation criteria (both in the electromagnetic calorimeter and hadronic calorimeter) are used to suppress the multijet background contribution from jets faking electrons. Events are retained if at least one isolated lepton with $p_{\text{T}} > 20\,\text{GeV}/c$ (referred to as the *first leg*) is reconstructed.

The selected events populate the W+jets sample. If a second reconstructed same flavor lepton with $p_{\text{T}} > 10\,\text{GeV}/c$ is found that forms an invariant mass in the range $(60 < m_{\ell\ell} < 110)\,\text{GeV}/\text{c}^2$ with the first leg, it is referred to as the *second leg* and the event is moved to the Z + jets sample. We require the selected events not to have additional same flavor leptons passing the tight first leg selection, removing background events with multiple candidates. While the Z sample is pure and characterized by the invariant mass $m_{\ell\ell}$ [9], the W sample is largely polluted by multijet backgrounds not containing a real W boson, as well as $\text{t} \to \text{Wb}$ decays. To identify a W boson candidate we form its transverse mass, $m_{\text{T}} = \sqrt{2p_{\text{T}}^{\ell}p_{\text{T}}^{\nu}(1-\cos\phi_{\ell\nu})}$, where p_{T}^{ℓ} is the lepton p_{T}, p_{T}^{ν} is represented by the transverse calorimetric energy imbalance (MET) computed from the calorimeter depositions and $\phi_{\ell\nu}$ is the angle between the lepton transverse momentum and the missing transverse momentum; we require $\text{MET} > 15$ GeV and $m_{\text{T}} > 30(20)\,\text{GeV}/c^2$ in the electron (muon) case.

4.3 Jet clustering

The event selection proceeds with the counting of the associated jets in the event. The expectation (validated by the results presented here) is that any jet definition can be used to construct the double ratio without altering the analysis strategy; the exception would be jets that are so inclusive that the first few jet clusterings use up all the available phase space, as discussed in reference [20].

We consider two scenarios based on the expected understanding of detector effects on the jet clustering and jet counting: i) at the LHC startup we consider the calorimetric response as not yet fully understood. In this scenario we use calo-jets and track-jets [21] reconstructed from calorimeter deposits and tracks, respectively, using the Seedless Infrared Safe Cone (SISCone) jet algorithm [22] with a cone size $R_{\text{cone}} = 0.5$ in the $(\eta \times \phi)$ space. The two sets of jets probe different parts of the phase space and are independent in terms of detector effects. ii) The second scenario assumes enough understanding of the detector to allow fully corrected calo-jets and PF-jets; this scenario would enable a quantitative direct comparison with parton-level QCD predictions (as they become available).

Events are selected with one or more calo-jets (track-jets) within $|\eta| < 3.0$ ($|\eta| < 2.4$) and $p_{\text{T}} > 30\,\text{GeV}/c$ ($p_{\text{T}} > 15\,\text{GeV}/c$). Track-jets are reconstructed from tracks with $|\eta| < 2.4$

consistent with the event primary vertex. PF-jets are clustered within $|\eta| < 3$ with best performance within $|\eta| < 2.4$. The detailed description of particle flow jet reconstruction at CMS can be found elsewhere [13]. The leptons from the VB candidate in the event are not considered as jets.

In the double ratio, systematic errors due to the mapping from partons to jets, the parton distribution functions, and other corrections substantially cancel [11]. Given the CMS high-precision silicon tracker that offers a very good momentum resolution, track-jets (and eventually PF-jets) can probe a part of the phase space where the calorimeter response is low and provide higher statistics VB + jets samples, despite the limited η acceptance compared to calo-jets. Table 1 summarizes the expected signal and background yields for $Z(\rightarrow ee)$ + jets and $W(\rightarrow e\nu)$ + jets and $Z(\rightarrow \mu\mu)$+jets and $W(\rightarrow \mu\nu)$+jets selection for calo-jet and track-jet counting for 100 pb^{-1} of integrated luminosity. The quoted errors are statistical only, related to the size of the available Monte Carlo datasets.

4.4 Signal yields

The detailed description of the extraction of the Z + jets signal events for each jet multiplicity bin is given in a dedicated Z + jets study [9]. To determine the number of W + jets events for each jet multiplicity bin, we perform a one-dimensional unbinned and extended ML fit to the m_{T}.

The W transverse mass m_{T} discriminates between events with a real W (signal and top) from the fake-W background events (dominated by multijets and other VB + jets events).

We discriminate between W + jets signal events and top events on the basis of an event heavy-flavour (hf) enrichment criterion. We split the selected sample into hf-enriched (top-like) and hf-depleted (W+jets, signal-like) sub-samples. The simple longitudinal and transverse event impact parameters D_z^{evt} and D_{xy}^{evt}, defined as the event's largest jet impact parameters, and computed using CMS standard impact parameter algorithms on the tracks matched to a jet, are used as an elementary yet reliable heavy-flavour enrichment measures.

Jets originating from light quarks (hf quarks) have small (large) impact parameter values. In the case of $W(\rightarrow \mu\nu)$ + jets the hf-depleted (*i.e.*, signal-enriched) sub-sample is defined requiring $D_z^{\mathrm{evt}} < 100\ \mu$m and $D_{xy}^{\mathrm{evt}} < 100\ \mu$m for both calo-jets and track-jets. In the case of $W(\rightarrow e\nu)$+ jets we require $D_{xy}^{\mathrm{evt}} < 180\ \mu$m and $D_z^{\mathrm{evt}} < 180\ \mu$m ($D_{xy}^{\mathrm{evt}} < 80\ \mu$m and $D_z^{\mathrm{evt}} < 80\ \mu$m) for calo-jets (track-jets)[2]. The breakdown of the signal and background events into the hf-enriched and hf-depleted datasets is encoded in a discrete variable θ_{hf}, set to one (zero) if the event belongs to the hf-enriched (hf-depleted) dataset. We then perform a fit on the transverse mass simultaneously in the hf-enriched (top and heavy-flavour multijets) and hf-depleted (W and light flavour multijets) sub-samples.

The likelihood used for the fit is written as a function of θ_{hf} and m_{T}:

$$\begin{aligned}\mathcal{L}_W &= \frac{e^{-(N_S+N_B+N_t)}}{(N_S+N_B+N_t)!}\prod_i\{[\epsilon_S\cdot N_S\cdot P_S(m_T^i)+\\ &\quad \epsilon_B\cdot N_B\cdot P_B(m_T^i)+\epsilon_t\cdot N_t\cdot P_t(m_T^i)](1-\theta_{hf})+\\ &\quad [(1-\epsilon_S)\cdot N_S\cdot P_S(m_T^i)+(1-\epsilon_B)\cdot N_B\cdot P_B(m_T^i)+\\ &\quad (1-\epsilon_t)\cdot N_t\cdot P_t(m_T^i)]\theta_{hf}\}\ , \qquad (1)\end{aligned}$$

[2]The definition of the hf-enriched and a hf-depleted subsamples has been optimized with Monte Carlo pseudo-experiments minimizing the statistical error on $W + \geq 3$ jets.

$W(\rightarrow e\nu)$ + jets Selection					
	W + jets	other VB + jets	$t\bar{t}$ + jets	Single t	QCD multijets
≥ 1 calo-jets (track-jets)	21897 ± 95 (61960 ± 161)	2285 ± 26 (3510 ± 32)	1823 ± 8 (1934 ± 8)	741 ± 5 (891 ± 5)	1247 ± 106 (5473 ± 244)
≥ 2 calo-jets (track-jets)	2922 ± 35 (10265 ± 65)	411 ± 11 (951 ± 17)	1245 ± 6 (1676 ± 7)	292 ± 3 (438 ± 4)	329 ± 50 (1245 ± 110)
≥ 3 calo-jets (track-jets)	389 ± 13 (1936 ± 28)	61 ± 4 (203 ± 8)	520 ± 4 (1101 ± 6)	54 ± 1 (140 ± 2)	44 ± 15 (243 ± 43)
≥ 4 calo-jets (track-jets)	54 ± 5 (363 ± 12)	8 ± 2 (42 ± 3)	132 ± 2 (490 ± 4)	9 ± 1 (34 ± 1)	9 ± 5 (46 ± 12)
$Z \rightarrow$ ee + jets Selection					
	Z + jets	other VB + jets	$t\bar{t}$ + jets	Single t	QCD multijets
≥ 1 calo-jets (track-jets)	1861 ± 23 (4525 ± 36)	8 ± 2 (15 ± 2)	44 ± 1 (50 ± 1)	4 ± 1 (5 ± 1)	6 ± 4 (6 ± 4)
≥ 2 calo-jets (track-jets)	260 ± 9 (794 ± 15)	1 ± 1 (2 ± 1)	20 ± 1 (33 ± 1)	1 ± 1 (2 ± 1)	3 ± 3 ($-$)
≥ 3 calo-jets (track-jets)	37 ± 3 (151 ± 7)	$-$ (1 ± 1)	5 ± 1 (13 ± 1)	$-$ ($-$)	$-$ ($-$)
≥ 4 calo-jets (track-jets)	3 ± 1 (29 ± 3)	$-$ ($-$)	1 ± 1 (4 ± 1)	$-$ ($-$)	$-$ ($-$)
$W(\rightarrow \mu\nu)$+jets Selection					
	W + jets	other VB + jets	$t\bar{t}$+ jets	Single t	QCD multijets
≥ 1 calo-jets (track-jets)	19331 ± 88 (66995 ± 163)	1409 ± 22 (5047 ± 43)	2317 ± 10 (2491 ± 10)	736 ± 5 (952 ± 6)	1790 ± 79 (12036 ± 205)
≥ 2 calo-jets (track-jets)	2685 ± 33 (11325 ± 67)	206 ± 9 (924 ± 18)	1531 ± 8 (2230 ± 10)	297 ± 3 (503 ± 5)	265 ± 30 (1856 ± 80)
≥ 3 calo-jets (track-jets)	356 ± 12 (2193 ± 30)	25 ± 3 (183 ± 8)	631 ± 5 (1530 ± 8)	60 ± 1 (175 ± 3)	14 ± 7 (306 ± 33)
≥ 4 calo-jets (track-jets)	51 ± 5 (404 ± 13)	4 ± 2 (30 ± 3)	162 ± 3 (683 ± 5)	10 ± 1 (49 ± 1)	$-$ (28 ± 10)
$Z(\rightarrow \mu\mu)$ + jets Selection					
	Z + jets	other VB + jets	$t\bar{t}$+ jets	Single t	QCD multijets
≥ 1 calo-jets (track-jets)	2351 ± 27 (7340 ± 47)	120 ± 7 (139 ± 8)	155 ± 2 (175 ± 3)	21 ± 1 (27 ± 1)	35 ± 11 (45 ± 13)
≥ 2 calo-jets (track-jets)	308 ± 10 (1328 ± 20)	31 ± 4 (53 ± 5)	90 ± 2 (140 ± 2)	9 ± 1 (16 ± 1)	7 ± 5 (14 ± 7)
≥ 3 calo-jets (track-jets)	36 ± 3 (242 ± 9)	7 ± 2 (18 ± 3)	32 ± 1 (81 ± 2)	2 ± 1 (6 ± 1)	3 ± 3 (7 ± 5)
≥ 4 calo-jets (track-jets)	6 ± 1 (44 ± 4)	2 ± 1 (2 ± 1)	8 ± 1 (32 ± 1)	$-$ (2 ± 1)	$-$

Table 1: Expected signal and background event yields in 100 pb^{-1} at $\sqrt{s} = 10$ TeV for for $Z \rightarrow$ ee + jets and $W(\rightarrow e\nu)$ + jets as well as $Z \rightarrow \mu\mu$ + jets and $W(\rightarrow \mu\nu)$ + jets selections. The QCD values are affected by large uncertainties due to limited Monte Carlo statistics.

where ϵ_X is the fraction of events for each component (Signal, Background, and top) belonging to the hf-depleted (signal enriched) subsample.

The value of ϵ_B is determined in the fit together with the three yields N_S, N_B, and N_t, since the shape of m_{T} allows one to disentangle this component from the rest in both the hf-enriched and hf-depleted samples. The efficiencies ϵ_S and ϵ_t are inputs to the fit which could be taken from Monte Carlo simulation. Much preferred is a fully data-driven method, measuring these efficiencies with data for the top and signal components. This requires the development of appropriate data control samples, as described below.

4.4.1 Data-driven extraction of ϵ_S and ϵ_t

A $\mathrm{t\bar{t}}$ control sample is selected using the same trigger, Z veto and muon p_{T} requirements as the $\mathrm{W}(\mu\nu)+$ jets signal selection. To isolate events coming from heavy flavour production, the impact parameter requirement of the W candidate muon is reversed. To better represent the top-like multijet topology we require ≥ 4 calo-jets. The sample contains a mixture of heavy flavour jets and jets from initial and final state radiation as well as top events. The event transverse and longitudinal impact parameter distributions are shown in Figure 1. Owing to the small dependence of the impact parameter variables on the inclusive jet multiplicity in $\mathrm{t\bar{t}}$ events, the agreement with the control sample persists as a function of jet multiplicity for up to ≥ 4 calo-jets and track-jets. Thus the correct ϵ_t input to the fit can be extracted directly from the control sample. The ϵ_t as extracted from this control sample agrees to within 1% with the one determined from the $\mathrm{t\bar{t}}$ Monte Carlo simulation. We use the $\mathrm{Z}(\rightarrow \ell\ell)$+jets candle data sample [9] to

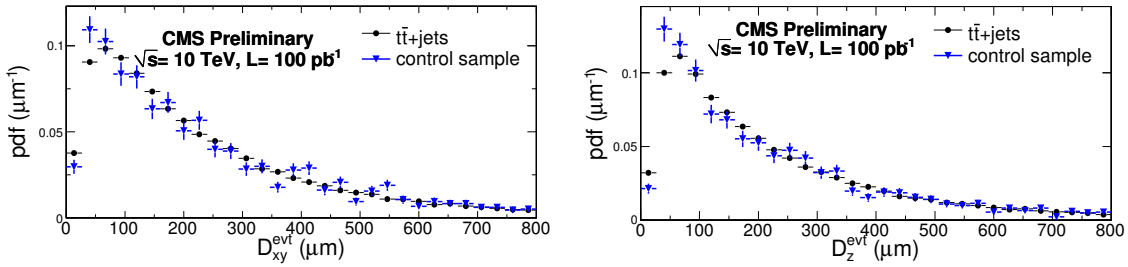

Figure 1: Event transverse (left) and longitudinal (right) impact parameter distributions in the $\mathrm{t\bar{t}}$ (dots) and heavy-flavour enriched (triangles) data control sample. The distributions are normalized to unitary area while the error bars correspond to 100 pb^{-1}. The same agreement is found for higher calo- and track-jet multiplicities.

model D_{xy}^{evt}, D_z^{evt} and measure ϵ_S for the signal $\mathrm{W}(\ell\nu)$+jets events directly in the data. W and Z production result in nearly identical final state topologies with respect to the D_{xy}^{evt} and D_z^{evt} variables, as shown in Figure 2, for example, in the muon channel with track-jet counting. W and Z production also exhibit nearly identical mixtures of heavy-flavour jets and light flavours from multijet associated production. Hence the pure $\mathrm{Z}+$ jets candle sample provides the ideal control sample for extracting ϵ_S. The value of ϵ_S as extracted from the candle control sample agrees to within 1% with the one determined from the $\mathrm{W}+$ jets Monte Carlo simulation.

4.4.2 Sophisticated b-tagging techniques

Use of the sophisticated b-jet tagging algorithms developed in CMS could produce a signal-enriched sub-sample with further reduced top background contribution, compared to what is obtained with the more primitive transverse and longitudinal jet impact parameters described above. Past experience at the Tevatron indicates that it takes time to understand a complex

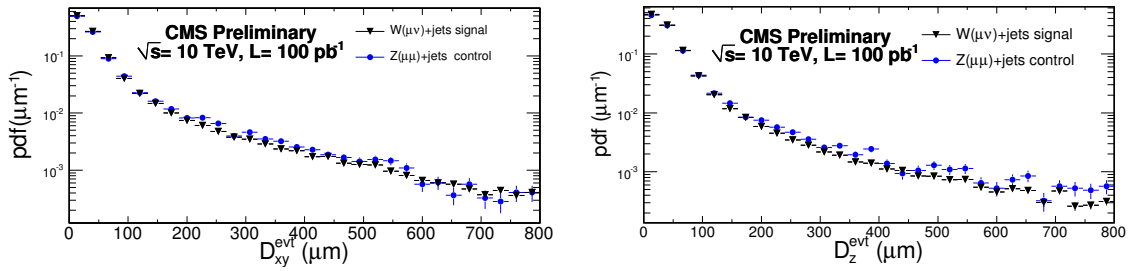

Figure 2: Event transverse (left) and longitudinal (right) impact parameter distributions in the Z + jets candle data sample and the signal W + jets sample. The distributions are normalized to unitary area while the error bars correspond to 100 pb^{-1}. The same agreement is found for higher calo- and track-jet multiplicities.

tracking system in a hadronic environment to the level required to master very sophisticated b-tagging algorithms. In addition B hadron production and decay models need to be improved and tuned with first LHC results. To study and compare the results in a mature data-taking era we do repeat the full analysis strategy using one of the CMS developed b-tagging methods. We replace the D_{xy}^{evt} and D_{z}^{evt} variables with an event variable defined as the maximum value of the output of a CMS b-tagger for all the considered jets. This new event variable is the 3D impact parameter significance constructed using a multivariate combination of the track's longitudinal and transverse impact parameters and their resolutions ($\text{MVA}_{3D}^{\text{evt}}$, [23]). We again optimize the selection of the b-tagger output as in the nominal analysis to define the b-enriched and b-depleted subsamples.

4.4.3 Signal and background parameterizations

The signal distribution of the dilepton invariant mass is described by a Gaussian-like function with asymmetric widths and non-Gaussian tails:

$$f(x; m, \sigma_L, \sigma_R, \alpha_L, \alpha_R) = N_s \cdot e^{-\frac{(x-m)^2}{2\sigma^2+\alpha(x-m)^2}}, \tag{2}$$

where $\sigma = \sigma_L(\sigma_R)$ for $x < m(x > m)$ and $\alpha = \alpha_L(\alpha_R)$ for $x < m(x > m)$. The background probability density function (PDF) is expected to be well modelled by either an exponential or a second-order polynomial as described in reference [9]. In the W + jets m_{T} fit a hybrid Crystal Ball function is used to describe the transverse mass distribution that incorporates the function of Equation (2) for the description of the left tail as follows:

$$f_{CC}(x; \alpha_{Cr}, \alpha_{CB}, n, m, \sigma) = N \cdot \begin{cases} \exp -\frac{(x-m)^2}{2\sigma^2+\alpha_{Cr}(x-m)^2} & \text{for } x-m<0 \\ \exp(-\frac{(x-m)^2}{2\sigma^2}), & \text{for } x-m\geq 0 \text{ and } \frac{x-m}{\sigma} > -\alpha_{CB} \\ A\cdot(B-\frac{x-m}{\sigma})^{-n}, & \text{for } x-m\geq 0 \text{ and } \frac{x-m}{\sigma} \leq -\alpha_{CB}\,, \end{cases} \tag{3}$$

where

$$A = \left(\frac{n}{|\alpha_{CB}|}\right)^n \cdot \exp\left(-\frac{|\alpha_{CB}|^2}{2}\right), \quad B = \frac{n}{|\alpha_{CB}|} - |\alpha_{CB}|\,. \tag{4}$$

The same values of σ and m are used in the two branches of Equation (3) hence the PDF is described by five independent parameters.

In the W + jets fit we distinguish the top background component ($t\bar{t}$ and single top) from the *other* backgrounds, which are dominated by the multijet contribution. In both cases the m_T distribution is well described using the same function for the signal component. We verify that the top distributions become almost identical to the signal when at least three jets are found.

The m_T line-shape of the signal and the top events can be studied using the Z + jets candle data sample. By ignoring one of the two leptons in the final states we compute m_T^Z as in a W + jets event and derive the corrections to apply to the W + jets and top Monte Carlo events. The line-shape of the fit variable for the multijet background is studied in the 'anti-muon' and 'anti-electron' multijet control samples. These are obtained by inverting the tracking muon isolation requirement and muon transverse impact parameter in the muon-enriched multijet sample, and all the electron isolation criteria and electron transverse impact parameter in the electron-enriched multijet sample. The rest of the W + jets selection path is kept unaltered. The m_T distribution for the multijet background is shown in Figure 3 for the events selected by the analysis and those in the anti-lepton control samples. The anti-lepton sample provides the control data sample for the validation of the analytical function describing the multijet background in the fit.

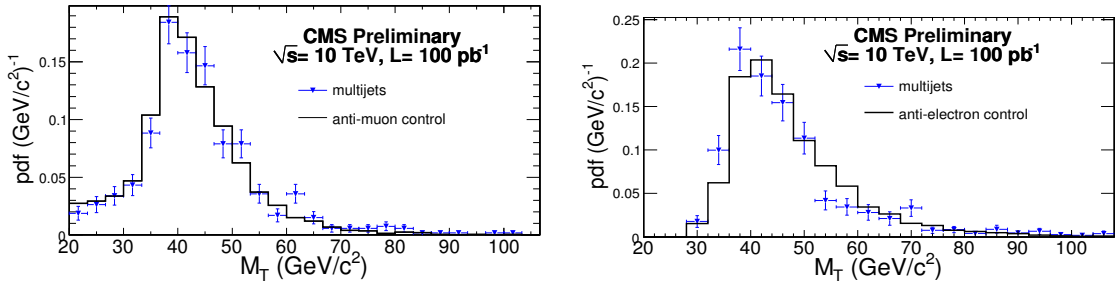

Figure 3: The transverse mass m_T distribution with muons (left) and electrons (right) in the lepton enriched multijet sample (points) and the control 'anti-lepton' multijet sample (histogram).

4.4.4 Fit results and tests

By performing a set of pseudo-experiments for each jet multiplicity, we estimate the expected statistical error on the signal yield for 100 pb^{-1}. The fits are performed on Monte Carlo samples generated from the distributions obtained from the full simulation. This allows one to perform the fit with unweighted events and to properly compute the statistical error of the fit result. These Monte Carlo tests demonstrate that the ML fits (for W + jets and Z + jets and for any jet multiplicity) are unbiased and that the 68% confidence interval computed using the likelihood ratio correctly covers the true number of events.

We perform another set of pseudo-experiments to estimate any bias induced on the fit by using exactly the same PDF for the W + jets signal and the top background component including the single top. We observe a large bias on the yield of the top events with no appreciable effect on the W + jets signal yields. This indicates a cross-feed of events between the multijet and top background components that has no impact on the results of this study. Figures 4 and 5 show the result of the m_T fit for the muon+jets and electron+jets final states with track-jet counting. The corresponding fit results of the Z+jets $m_{\ell\ell}$ ML fit are similar to the ones presented in [9].

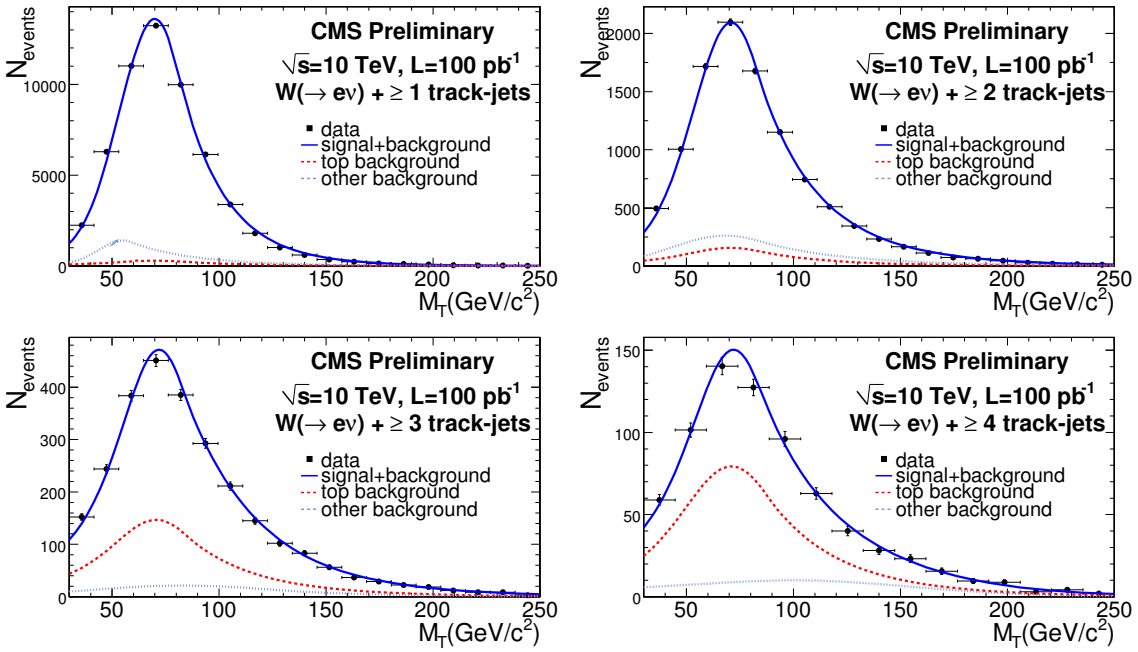

Figure 4: Projection of the likelihood at maximum on m_{T} for $\mathrm{W}(\rightarrow \mathrm{e}\nu) + \geq 1$ track-jets (top left), $\mathrm{W}(\rightarrow \mathrm{e}\nu) + \geq 2$ track-jets (top right), $\mathrm{W}(\rightarrow \mathrm{e}\nu) + \geq 3$ track-jets (bottom left), and $\mathrm{W}(\rightarrow \mathrm{e}\nu) + \geq 4$ track-jets (bottom right). The 'data' sample corresponding to 100 pb^{-1} statistics is overlaid. The error bars correspond to the expected precision. The dashed line represents the top background and the dotted line the other background contributions.

4.5 The W + ≥ n jets to Z + ≥ n jets ratio

With the ML fit to the four different jet multiplicity samples we measure the yields of Z + n and W + n jets as a function of jet multiplicity. Defining C_{VB} as the VB + n jets to VB + $(n+1)$ jets yield ratio, we expect C_{VB} to be independent of n, within errors. Under the assumption that C_{VB} is a constant, the ratio of inclusive VB + n jets ($\geq n$ jets) to inclusive VB + $(n+1)$ jets ($\geq n+1$) is identical to the ratio of exclusive VB + n jets ($= n$) to exclusive VB + $(n+1)$ jets ($=n+1$). Thus physically C_{VB} represents the cost of adding an extra jet to VB + n jet production at some fixed order in α_s. The extracted value of C_{VB} depends on the jet definition: e.g., increasing the jet p_{T} threshold for a fixed cone size increases C_{VB}, while decreasing the cone size for a fixed jet p_{T} threshold also increases C_{VB}. Indeed, the difference in the C_{VB} values extracted from the calo-jet counting versus track-jet counting is largely due to the fact that track-jets probe a lower p_{T} region of the phase space. By using both track-jets and calo-jets counting, the prediction of a constant C_{VB} can be verified in different regions of the phase space and using independent detector elements. Additionally, by using corrected calo-jets or PF-jets a detailed quantitative comparison with the parton-level QCD predictions could eventually be made.

The fit of the measured yields to an exponential, shown in Figure 6 for the electron and muon channels with track-jet and calo-jet counting, confirms the validity of the constant ratio and double ratio assumption returning fit probabilities between 65% and 99%. Here the errors reflect the expected statistical precision on data, as estimated from pseudo-experiment Monte Carlo tests without taking into account the statistical correlation between the successive bins. We expect a similar picture to emerge from the first LHC data.

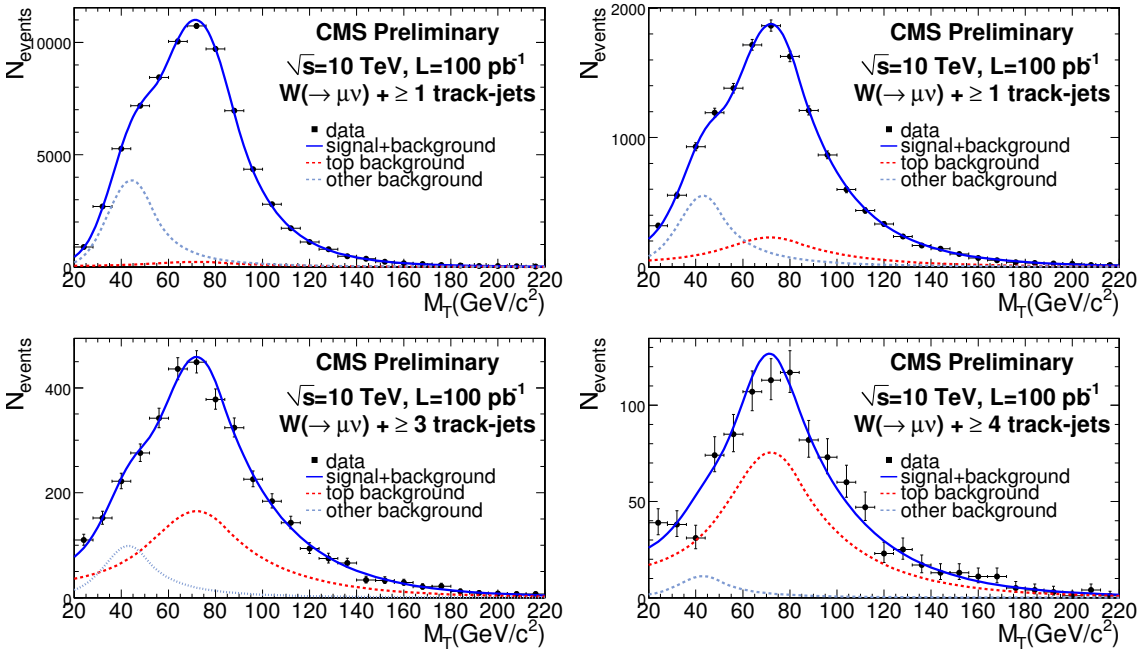

Figure 5: Projection of the likelihood at maximum on m_{T} for $\mathrm{W}(\rightarrow \mu\nu) + \geq 1$ track-jets (top left), $\mathrm{W}(\rightarrow \mu\nu) + \geq 2$ track-jets (top right), $\mathrm{W}(\rightarrow \mu\nu) + \geq 3$ track-jets (bottom left), and $\mathrm{W}(\rightarrow \mu\nu) + \geq 4$ track-jets (bottom right). The 'data' sample corresponding to 100 pb^{-1} statistics is overlaid. The error bars correspond to the expected precision. The dashed line represents the top background and the dotted line the other background contributions.

The value of C_{VB} for calo-jets corresponds to the value obtained for generator-level jets, in the same rapidity range, for p_{T} threshold $\sim$60 GeV/c in agreement with the expected calorimeter response [9]. With understood data, the slopes and double ratios as extracted from corrected calo-jets and PF-jets could be directly compared to QCD predictions, represented here by the generator-level jets from leading order QCD Monte Carlo with jet-parton matching. We validate that this is the case taking (as an example in the electron channel) 60 GeV/c p_{T} threshold for both corrected calo-jets and PF-jets. The results are shown in Figure 7 and Table 2. The table includes also the values of the C_{VB}'s corrected for the selection efficiency within each jet multiplicity bin (denoted C'_{VB}'s). We find that C_{VB} is constant for all types of jet counting and for both the W + jets electron selection and Z + jets dielectron selection. Further we find that the double ratio $C_{\mathrm{W}}/C_{\mathrm{Z}} \equiv \frac{\mathrm{W} + n \text{ jets} \ / \ \mathrm{W} + (\mathrm{n}+1) \text{ jets}}{\mathrm{Z} + n \text{ jets} \ / \ \mathrm{Z} + (\ n+1) \text{ jets}}$ is consistent with 1 within the precision obtained independent of the jet definition. The corresponding results in the muon channel for nominal calo-jet and track-jet counting are shown in Table 3. In Figure 8 the ratio of the yields as a function of inclusive track-jet multiplicity is shown in the muon channel with ('Eff. corrected') and without ('Raw yields') the selection efficiency correction within each jet multiplicity bin. In all cases the jet counting is inclusive ($\geq n$) resulting in a statistical correlation between the successive bins. The efficiency dependence on the jet multiplicity cancels out in the ratio of the slopes for W + jets and Z + jets. As shown in the tables, the ratios of the slopes are compatible with 1 within errors, for both electron and muon final states. In particular, the track-jet selection yields a high statistics sample that shows the overall consistency of the analysis with relatively high precision. This demonstrates that detector effects related to the event selection (both the leptonic selection and the associated jet selection) as well as any interference of the two due to detector or reconstruction effects (e.g. MET and isolation criteria)

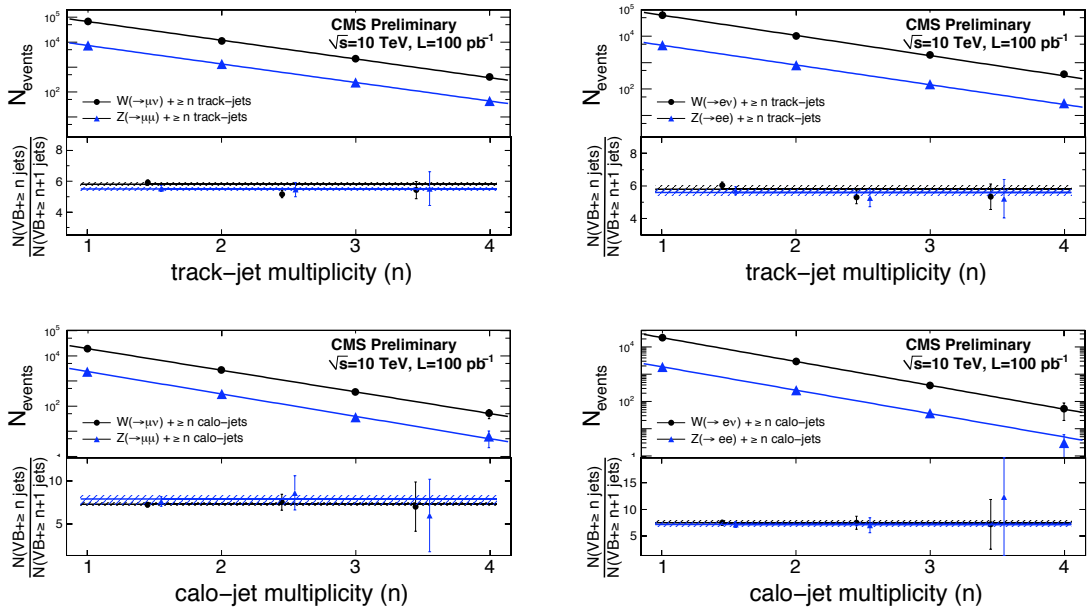

Figure 6: The (dN/dn_{jet}) distributions and exponential fit for $W(\rightarrow \mu\nu)$ + jets and $Z(\rightarrow \mu\mu)$ + jets (left) and $W(\rightarrow e\nu)$ + jets and $Z(\rightarrow ee)$ + jets (right) and $W(\rightarrow \mu\nu)$ + jets and $Z(\rightarrow \mu\mu)$ + jets (right) for track-jet counting (top) and calo-jet counting (bottom). The resulting constant C_{VB} ratio values of $Z + n$ jets to $Z + (n+1)$ jets and $W + n$ jets to $W + (n+1)$ summarized in Table 2 for electron final states and Table 3 for muon final states, are also shown.

	calo-jet	track-jet	calo$^{cor}-jet$	PF-jet
C_W	7.5±0.4	5.8±0.1	7.1±0.3	7.0±0.3
C_Z	7.2±0.4	5.6±0.2	6.9±0.4	7.0±0.4
C_W/C_Z	1.04±0.08	1.04±0.04	1.03±0.07	1.01±0.08
C'_W	6.9±0.3	5.5±0.1	6.6±0.3	6.5± 0.3
C'_Z	7.05±0.4	5.4±0.2	6.7±0.4	6.9±0.4
C'_W/C'_Z	0.98±0.07	1.01±0.04	0.98±0.07	0.95±0.07

Table 2: C_{VB} ratios and C_W/C_Z double ratios in the electron channel for calo-(uncorrected 30 GeV/c), track-(15 GeV/c), corrected calo-(60 GeV/c) and PF-(60 GeV/c) jet counting. The C'_{VB} values are computed taking into account the selection efficiency within each multiplicity bin. Within the precision achieved the ratios and double ratios are consistent with expectations.

mostly cancel out in the W + jets to Z + jets ratio. Projecting on 100 pb^{-1} of well-understood data beyond the startup we performed the analysis using one of the sophisticated b-tagging CMS algorithms in the case of the electron channel. We find a constant double ratio and an improvement on the uncertainty of the directly measured W+3(4) jets yields of 20(30)%.

4.6 Prediction of W + ≥ 3,4-jets

Assuming that we find a constant ratio in the data, we would use the VB + ≥ 1 jet and VB + ≥ 2 jet yields to predict the yields for higher jet multiplicities. These yields would then be used to perform a test of the SM, and to provide a data-based estimate of the VB+jets backgrounds to other SM processes as well as searches for new physics. The results expected for 100 pb^{-1} are illustrated in Table 4 and Table 5 for events with W+ ≥ 3 jets and W + ≥ 4 jets

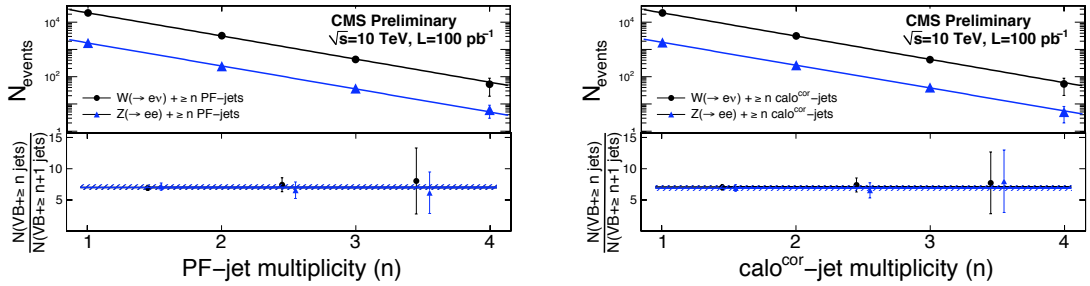

Figure 7: The $(dN/d\mathrm{n_{jet}})$ distributions and exponential fit for $\mathrm{W}(\rightarrow \mathrm{e}\nu)$ + jets (left) using PF-jets and (right) corrected jets. The resulting constant C_{VB} ratio values of $\mathrm{Z} + n$ jets to $\mathrm{Z} + (n+1)$ jets and $\mathrm{W} + n$ jets to $\mathrm{W} + (n+1)$ summarized in Table 2 for electron final states, are also shown.

	calo-jet	track-jet
C_W	7.3±0.2	5.8±0.1
C_Z	7.8±0.5	5.5±0.1
C_W/C_Z	0.93±0.06	1.04±0.03
C'_W	6.9±0.2	5.5±0.1
C'_Z	7.2±0.5	5.2±0.1
C'_W/C'_Z	0.95±0.07	1.05±0.03

Table 3: C_{VB} ratios and C_W/C_Z double ratios in the muon channel for nominal calo-jet and track-jet counting. The C'_{VB} values are computed taking into account the selection efficiency within each multiplicity bin. Within the precision achieved the ratios and double ratios are consistent with expectations.

respectively. Figure 9 illustrates two sets of comparisons based on Tables 4 and 5. The results shown use as inputs: i) the $\mathrm{W} + \geq 1$ jet and $\mathrm{W} + \geq 2$ jet yields, ii) the value of the ratio C_Z defined earlier which is obtained from the tight selection of the Z + jets, and iii) the value of the ratio $r = \frac{N(\mathrm{W} + \geq 1 \text{ jet})}{N(\mathrm{Z} + \geq 1 \text{ jet})}$. As shown in the table, the uncertainty on $\mathrm{W} + \geq n$ jets from the prediction is better than the direct measurement, and the contributions of systematic errors are largely suppressed.

5 Conclusions

We have presented a simulation study of the $\mathrm{W} + n$ jets over $\mathrm{Z} + n$ jets ratio for $n \geq 1$ at $\sqrt{s} = 10$ TeV at the Compact Muon Solenoid with 100 pb^{-1} of data. In the ratio, systematic uncertainties that grow rapidly with n, being dominated by uncertainties in the identification of jets and the jet energy scale, cancel, making this study viable at the LHC startup. The cancellation of systematic uncertainties is predominantly due to the correlation in the jet counting uncertainties in the numerator and denominator. Other systematic uncertainties associated with the luminosity, parton distribution functions, detector acceptance and efficiencies, are also expected to substantially cancel ; this allows both a greater sensitivity to new physics contributions in these channels and an accurate means of prediction for the VB + n jets backgrounds to SM processes as well as NP searches. As an example we present the prediction of the $\mathrm{W} + \geq 3, 4$ jets yields. Given the large contribution of top quark related processes for $\geq 3, 4$ jets, the ratio could also provide additional understanding of top-related measurements

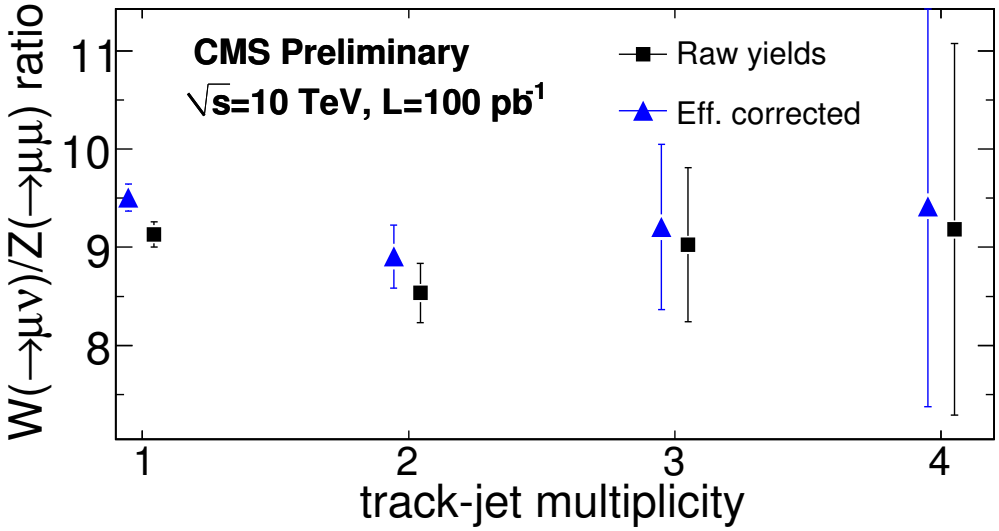

Figure 8: Ratio of W($\rightarrow \mu\nu$) + jets event yields to Z($\rightarrow \mu\mu$) + jets event yields as a function of track-jet multiplicity; the result is also shown corrected for the efficiency within each jet multiplicity bin. Similar results are obtained for other jet definitions both in the electron and muon final states.

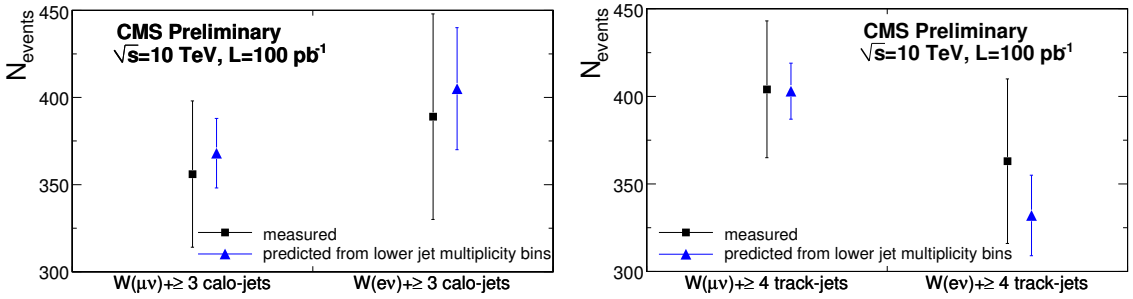

Figure 9: Comparison between the expected number of selected W + $\geq$3 calo-jets (left) and W + $\geq$4 track-jets (right) in 100 pb^{-1} and the prediction based on the yields with lower jet multiplicities and the Z + jets slope, C_Z, as measured in this analysis.

at the LHC startup.

6 Acknowledgements

We are grateful to Susan Leech O'Neale for editing the manuscript.

References

[1] **CDF** Collaboration, T. Aaltonen et al., "Measurement of the cross section for W-boson production in association with jets in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV," *Phys. Rev.* **D77** (2008) 011108, arXiv:0711.4044. doi:10.1103/PhysRevD.77.011108.

[2] **CDF - Run II** Collaboration, T. Aaltonen et al., "Measurement of Inclusive Jet Cross Sections in $Z/\gamma^*(\rightarrow ee)+$ jets Production in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV," *Phys. Rev.*

Sample	Measured	$N(\mathrm{W} + \geq 2\,\text{jets})/C_\mathrm{Z}$	$N(\mathrm{Z} + \geq 2\,\text{jets}) \cdot r/C_\mathrm{Z}$
Muon Channel			
$\mathrm{W}(\mu\nu) + \geq 3$ calo-jets	356 ± 42	368 ± 20	347 ± 29
$\mathrm{W}(\mu\nu) + \geq 3$ track-jets	2193 ± 83	2138 ± 50	2284 ± 91
Electron Channel			
$\mathrm{W}(\mathrm{e}\nu) + \geq 3$ calo-jets	389 ± 59	405 ± 35	424 ± 40
$\mathrm{W}(\mathrm{e}\nu) + \geq 3$ track-jets	1936 ± 131	1846 ± 82	1955 ± 106

Table 4: Comparison between the expected number of selected W $+ \geq 3$ jet events in 100 pb^{-1}, and the prediction based on the yields with lower jet multiplicities and the Z $+$ jets slope, C_Z, measured in this analysis.

Sample	Measured	$N(\mathrm{W} + \geq 2\,\text{jets})/C_\mathrm{Z}^2$	$N(\mathrm{Z} + \geq 2\,\text{jets}) \cdot r/C_\mathrm{Z}^2$
Muon Channel			
$\mathrm{W}(\mu\nu) + \geq 4$ calo-jets	51 ± 20	50 ± 4	48 ± 5
$\mathrm{W}(\mu\nu) + \geq 4$ track-jets	404 ± 39	403 ± 16	431 ± 22
Electron Channel			
$\mathrm{W}(\mathrm{e}\nu) + \geq 4$ calo-jets	54 ± 34	56 ± 7	59 ± 8
$\mathrm{W}(\mathrm{e}\nu) + \geq 4$ track-jets	363 ± 47	332 ± 23	352 ± 27

Table 5: Comparison between the expected number of selected W $+ \geq 4$ jets in 100 pb^{-1} and the prediction based on the yields with lower jet multiplicities and the Z $+$ jets slope, C_Z, as measured in this analysis.

Lett. **100** (2008) 102001, arXiv:0711.3717.

[3] **D0** Collaboration, V. M. Abazov et al., "Measurement of the shape of the boson transverse momentum distribution in ppbar →Z/gamma* → ee+X events produced at sqrts=1.96 TeV," *Phys. Rev. Lett.* **100** (2008) 102002, arXiv:0712.0803. doi:10.1103/PhysRevLett.100.102002.

[4] J. M. Campbell and R. K. Ellis, "Next-to-leading order corrections to W + 2jet and Z + 2jet production at hadron colliders," *Phys. Rev.* **D65** (2002) 113007, arXiv:hep-ph/0202176. doi:10.1103/PhysRevD.65.113007.

[5] J. M. Campbell, R. K. Ellis, and D. L. Rainwater, "Next-to-leading order QCD predictions for W + 2jet and Z + 2jet production at the CERN LHC," *Phys. Rev.* **D68** (2003) 094021, arXiv:hep-ph/0308195. doi:10.1103/PhysRevD.68.094021.

[6] R. K. Ellis, K. Melnikov, and G. Zanderighi, "Generalized unitarity at work: first NLO QCD results for hadronic W^+ 3jet production," arXiv:0901.4101.

[7] C. F. Berger et al., "Precise Predictions for W + 3 Jet Production at Hadron Colliders," arXiv:0902.2760.

[8] R. K. Ellis, W. T. Giele, Z. Kunszt, K. Melnikov, and G. Zanderighi, "One-loop amplitudes for W+3 jet production in hadron collisions," arXiv:0810.2762.

[9] **CMS** Collaboration, G. L. Bayatian et al., "Study of the Z production in association with jets in proton-proton collisions at $\sqrt{s} = 10$ TeV with the CMS detector at the CERN LHC,". CMS-EWK-08-006-1 (2009).

[10] H. Baer, V. Barger, and G. Shaughnessy, "SUSY backgrounds to Standard Model calibration processes at the LHC," arXiv:0806.3745.

[11] E. Abouzaid and H. J. Frisch, "The Ratio of W + N jets to Z^0/γ^* + N jets versus N as a test precision test of the standard model," *Phys. Rev.* **D68** (2003) 033014, arXiv:hep-ph/0303088. doi:10.1103/PhysRevD.68.033014.

[12] **CMS** Collaboration, R. Adolphi et al., "The CMS experiment at the CERN LHC," *JINST* **0803** (2008) S08004. doi:10.1088/1748-0221/3/08/S08004.

[13] **CMS** Collaboration, G. L. Bayatian et al., "Particle Flow Event Reconstruction in CMS and Performance for Jets, Taus and MET,". PF-09-001 (2009).

[14] F. Maltoni and T. Stelzer, "MadEvent: Automatic event generation with MadGraph," *JHEP* **02** (2003) 027, arXiv:hep-ph/0208156.

[15] T. Sjostrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual," *JHEP* **05** (2006) 026, arXiv:hep-ph/0603175.

[16] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, "Matching matrix elements and shower evolution for top- quark production in hadronic collisions," *JHEP* **01** (2007) 013, arXiv:hep-ph/0611129.

[17] S. Kretzer, H. L. Lai, F. I. Olness, and W. K. Tung, "CTEQ6 parton distributions with heavy quark mass effects," *Phys. Rev.* **D69** (2004) 114005, arXiv:hep-ph/0307022. doi:10.1103/PhysRevD.69.114005.

[18] **CMS** Collaboration, G. L. Bayatian et al., "Measurement of the Zbb,$Z \to \ell\ell$ cross section with 100 pb^{-1} of Early CMS data at the LHC,". CMS-EWK-08-001 (2008).

[19] **CMS** Collaboration, G. L. Bayatian et al., "CMS physics: Technical design report,". CERN-LHCC-2006-001.

[20] F. A. Berends, H. Kuijf, B. Tausk, and W. T. Giele, "On the production of a W and jets at hadron colliders," *Nucl. Phys.* **B357** (1991) 32–64. doi:10.1016/0550-3213(91)90458-A.

[21] **CMS** Collaboration, P. Azzurri, "Jet Reconstruction with charged tracks only in CMS," arXiv:0901.1541.

[22] G. P. Salam and G. Soyez, "A practical Seedless Infrared-Safe Cone jet algorithm," *JHEP* **05** (2007) 086, arXiv:0704.0292.

[23] **CMS** Collaboration, G. L. Bayatian et al., "Algorithms for b Jet Identification in CMS,". CMS-BTV-08-001 (2008).